

## “Moneyball” Analysis and Identifying Big Data and Analytics Applications

### *Oakland A’s and Billy Beane*

The Oakland A’s started the 2002 baseball season analytically impaired. The general manager, Billy Beane, knew that given a budget a third the size of some teams, that they could not compete against their competitors with traditional tactics. What’s remarkable about the Oakland A’s and Beane, is that they embarked on the roadmap to becoming an analytical competitor before they even knew they were on it.

First, Beane identified the business problem correctly. While co-workers were telling Beane the problem was a lack of talent on the team, Beane knew the real problem was the small budget. This acknowledgement allowed Beane to focus on the right problem, such that when he was introduced to analytical techniques, he saw a great opportunity to identify undervalued players. At an early stage of the 2002 season, the A’s already possessed the focus and ambition needed to become an analytical competitor.

An important consideration of Beane’s success was the availability of data to mine. This is a rare occurrence in business, that all the data you need is already there sitting and waiting to be consumed. But in baseball, runs batted in, homeruns, years played, etc, are all widely available. And even more convenient for Beane is that he had both internal and external data. He had a mountain of data about all players in professional baseball. Without this information, they would have needed to invest a substantial effort in building that information, which they did not have the resources to do.

Furthermore, when Beane hired Peter Brand, he hired someone that knew how to ask the right questions of the data. He saw past all of the flashy statistics, to distill it to getting the player on a base. With this metric, they could mathematically arrive at undervalued players in professional baseball. This focus was another key component to their success, and was the moment that they achieved the stage of analytical aspirations. They had the technical support now, in addition to the support of management and the ambition. This gave them three of the four pillars of analytical competition<sup>1</sup>. They only lacked enterprise-wide analysis.

The reason they lacked enterprise-wise analysis, is because they did not share their knowledge with all of the right people to achieve success. It was not until later in the season, when Beane and Peter shared each player's analysis with the players, that they began to see a return on investment. By giving the entire organization insight into the analytics, the entire organization was finally able to act on that information. This is how the A's attained a 20 game winning streak in 2002, and an average \$260K per win versus the Yankee's \$1,400K per win<sup>2</sup>.

There were several things that Beane could have done to more effectively implement analytics with the A's roster. First, he took a 'full-steam' approach with his co-management, with little regard to sharing the details of his plan. If he spent more time explaining the details of his plan and sharing the success stories that Peter Brand had with the Cleveland Indians, he would have had more buy-in. Instead we saw more of a dictator giving orders. Second, it became obvious that Beane should have shared the predictions of the models with the players sooner. It gave players the opportunity to take advantage of strengths, and develop their weaknesses.

---

<sup>1</sup> Davenport, T.H., Harris, J.G.. (2007). Competing on analytics: the new science of winning. Harvard Business School Press.

<sup>2</sup> Miller, B. (Director). (2012). Moneyball [Motion picture]. USA: Sony Pictures Home Entertainment.

## *Airbnb*

Now, let's go across the bay and take a look at the analytical competitiveness of Airbnb. Airbnb has a record of using analytics for every aspect of their business. "We want to apply data to every decision. We want to be a very data-driven company.", says Airbnb's Vice President of Engineering, Mike Curtis.

<sup>3</sup> Given that Airbnb generates more than 20 terabytes of data to analyze every day, and employs more than 50 data scientists<sup>3</sup>, they are in a position to be an analytic competitor and data-driven company.

Airbnb has identified several distinctive capabilities to keep it competitive in the travel housing market. One such advantage is to be able to select features of where you are staying. Where many travel sites offer lists of hotels to stay at, Airbnb can offer customers open-air huts on the beach, or a jacuzzi in a mountain, or wifi in a small village. Because of this capability, they developed models to predict the most important features for customers. Their models work on the aggregate, but are also personally customized. If they notice you always want an apartment with A/C, those apartments become more relevant for you. They used these models to also determine that the most important feature of bookings is representative photographs. This lead Airbnb to offer professional photographers for free to some of its rental owners.<sup>2</sup>

Their search engine is another area where they've invested predictive analytics. They have created a customer-driven model, where customer bookings mapped to keywords power the search engine. Furthermore, they use analytics to identify fraudulent listings, and use demographics to optimize the layout of their homepage. For example, Airbnb determined that Asian markets were confused by

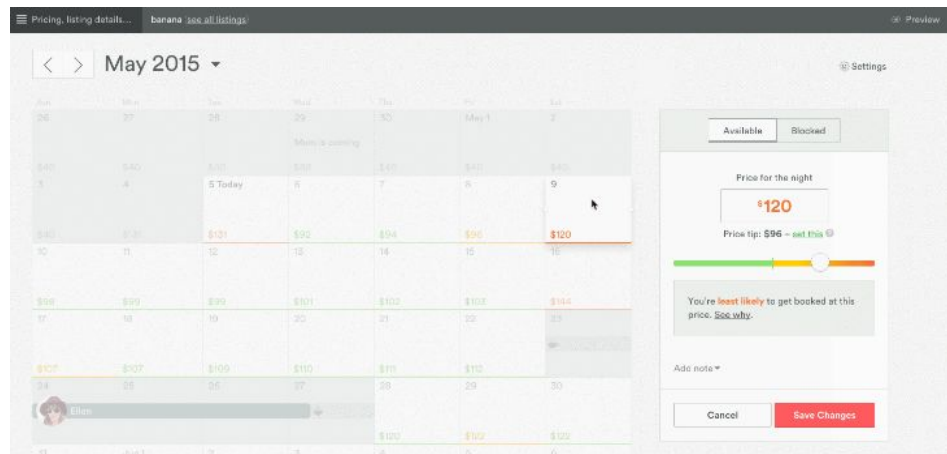
---

<sup>3</sup> Rigmenam, M.. (2015, February 12). AirBnB matches apartments, castles, and igloos with guests using big data. <https://datafloq.com/read/airbnb-matches-apartments-castles-igloos-guests-bi/295>

cluttered links on the homepage. By tweaking the homepage for Asian markets, they found a 10% conversion in bookings.<sup>4</sup>

One other impressive tool from Airbnb is their Price Tips feature, that analyzes hundreds of predictors to determine what price a rental is most likely to rent out for a given day. It looks at the day of the week, special events, comparable listings, etc., to arrive at costs that will get a renter maximum occupancy.

Furthermore, they distilled the complexity of the analysis into an easy-to-understand calendar of color-mapped prices, so that renters can make an informed decision on how to price their rentals.



Airbnb is also freely sharing their data efforts on their website for [Airbnb nerds](http://nerds.airbnb.com/data/)<sup>5</sup>. It proves the company's commitment to analytics and endorsing a culture of openness. While it's clear Airbnb is applying analytics in almost every aspect of its business, they could spend more thought on post-booking considerations. For example, if you book a cottage in the middle of Montana, what needs will the customer have in terms of getting to the cottage? That would be a possible deterrent for someone that could result in no booking. By investing more in solving those post-booking unknowns, Airbnb will be even more profitable.

<sup>4</sup> Song, M., Shan, C.. (2014, May 18). How airbnb used data to propel its growth to a \$10B valuation. <http://venturebeat.com/2014/05/18/how-airbnb-used-data-to-propel-its-growth-to-a-10b-valuation/>

<sup>5</sup> <http://nerds.airbnb.com/data/>